PSIGEN™

PSI:CAPTURE

OCR Module

Application User Guide

August 2009

PSIGEN Software, Inc., *8 Goodyear Suite 110*
Irvine, CA 92618 – USA

sales@psigen.com
www.psigen.com

About this Guide

This document is intended for the typical user of the PSI:Capture OCR Module for processing various document types. These document types should be configured by an administrator prior to the start of the capture process. Refer to the Administrative Guide as necessary for information relating to configuration of document types.

About PSIGEN Software, Inc.

PSIGEN Software is a fast growing, privately held, provider of document capture and retrieval software solutions. Our product lines address production capture, distributed capture, web based retrieval, including the Microsoft SharePoint platform. Our largest customers include scanning service providers, business process outsourcers, and the corporate services divisions of Fortune 500 corporations. We have significant customer concentration in financial services, healthcare, and manufacturing. Our channel partners include document imaging value added resellers, major brand MFP dealers, Microsoft solution providers, and document management software vendors.

Table of Contents

OCR Module Overview

OCR is an acronym for Optical Character Recognition. The core of the OCR module is a 'recognition' process: the OCR engine software attempts to read each and every letter/number/word on an image, and write it out to various file formats. The results of OCR depend on multiple aspects of image quality. Image quality generally stems from the paper quality, print type (font style), and print quality of the original. One of the most widely used file formats on the market today is the Adobe Portable Document Format (PDF). The OCR module has the ability to rapidly convert to and output PDF files, with the OCR as hidden text in each PDF. The results of the full text OCR module can be editable or non-editable files in a file format and directory of the user's choice. Supported OCR Output Types include: PDF (Image Only, Image w/ Hidden Text, Normal), HTML 4, Lotus 1 2 3, Lotus Ami Pro, MS Excel 97-03, MS Word 97-03, MS RTF, Text (Plain, Comma Delimited, Formatted, Tab Delimited, Line Breaks), and WordPerfect 8/9/10. The OCR result files are stored in the batch folder, and their existence has no effect on the TIFF images. These files are just another group of files created when this step is added to a capture workflow.

User Interface Overview

The PSI:Capture interface is divided into different functional areas, as seen in the layout below. These panes of the user interface will be referenced throughout this manual.
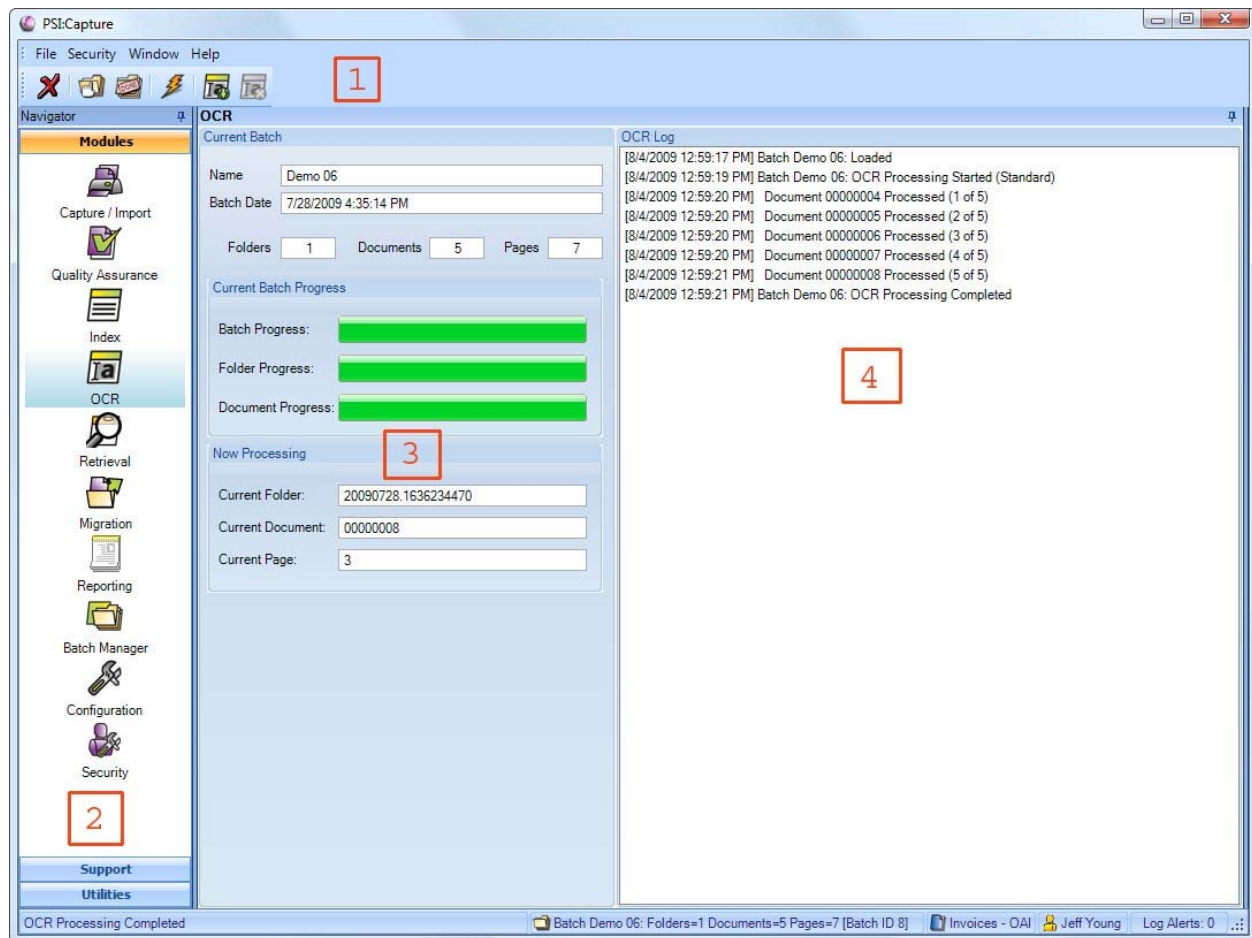
1. Toolbar Banner
2. Navigator Pane
3. Batch Progress Pane
4. OCR Log

Toolbar

OCR Toolbar

![Close Module icon] **Close Module** – Closes the current module.

![Select Batch icon] **Select Batch** – Opens the Batch Manager list for selection of existing batches for processing.

![Close Batch icon] **Close Batch** – Brings up the "Close Batch" dialog box.

![Toggle Auto Processing icon] **Toggle Auto Processing** – Enables and disables automatic OCR processing (on by default).
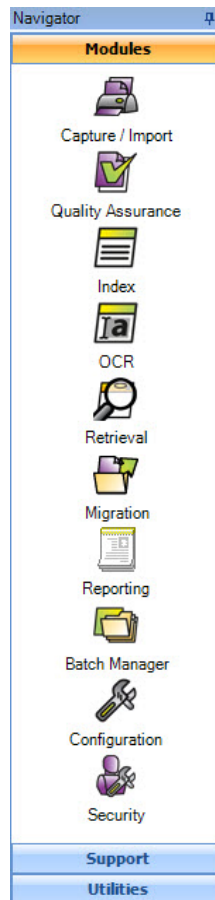
![Start OCR icon] **Start OCR** – Begins the OCR process.

![Stop OCR icon] **Stop OCR** – Stops the OCR process.

Navigator Pane

Navigator Pane can either be pinned to the desktop (default) or set to auto hide.

Modules Tab

Capture / Import – Opens the Capture/Import module.

Quality Assurance – Opens the Quality Assurance module.

Index – Opens the Index module.

OCR – Opens the OCR module.

Retrieval - Opens the Retrieval module.

Migration – Opens the Migration module.
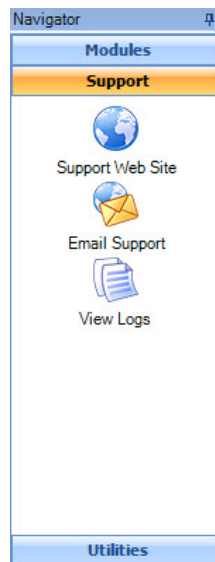
Reporting – Opens the Reporting module.

Batch Manager – Opens the Batch Manager module.

Configuration – Opens the Configuration module.

Security – Opens the Security module.

Support Tab

Contains links to support website, support email, and view logs.

Support Web Site

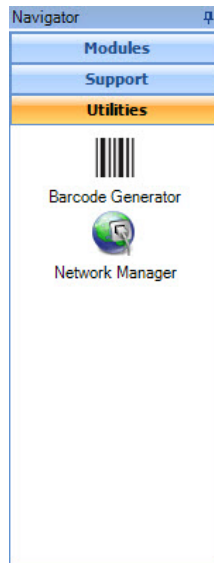Opens the default web browser to www.psigen.com/support

Email Support

Opens the default mail service for user to compose an email to support@psigen.com.

View Logs

Displays a list of PSI:Capture error and processing logs.

Utilities Tab

Contains Barcode Generator module and Network Manager.

Barcode Generator

Opens up a utility for users to create/save/print custom separator/indexing barcode sheets.

Network Manager

Allows user to manage the port and IP settings for PSIGEN Platform Services and License Server.

Batch Progress Pane

Current Batch

Displays the name and date of the current batch, as well as the folder/document/page count.

Current Batch Progress

Gives graphical representation of the OCR progress.

Now Processing

Indicates the current page OCR is processing.

OCR Log

A timestamp of when each batch and document is loaded, proccessed, and completed.

```
OCR Log
[8/4/2009 1:36:39 PM] Batch Demo 11: Loaded
[8/4/2009 1:37:04 PM] Batch Demo 11: OCR Processing Started (Standard)
[8/4/2009 1:37:05 PM]   Document 00000001 Processed (1 of 10)
[8/4/2009 1:37:05 PM]   Document 00000002 Processed (2 of 10)
[8/4/2009 1:37:05 PM]   Document 00000003 Processed (3 of 10)
[8/4/2009 1:37:05 PM]   Document 00000004 Processed (4 of 10)
[8/4/2009 1:37:06 PM]   Document 00000005 Processed (5 of 10)
[8/4/2009 1:37:06 PM]   Document 00000006 Processed (6 of 10)
[8/4/2009 1:37:06 PM]   Document 00000007 Processed (7 of 10)
[8/4/2009 1:37:06 PM]   Document 00000008 Processed (8 of 10)
[8/4/2009 1:37:06 PM]   Document 00000009 Processed (9 of 10)
[8/4/2009 1:37:06 PM]   Document 00000010 Processed (10 of 10)
[8/4/2009 1:37:06 PM] Batch Demo 11: OCR Processing Completed
```
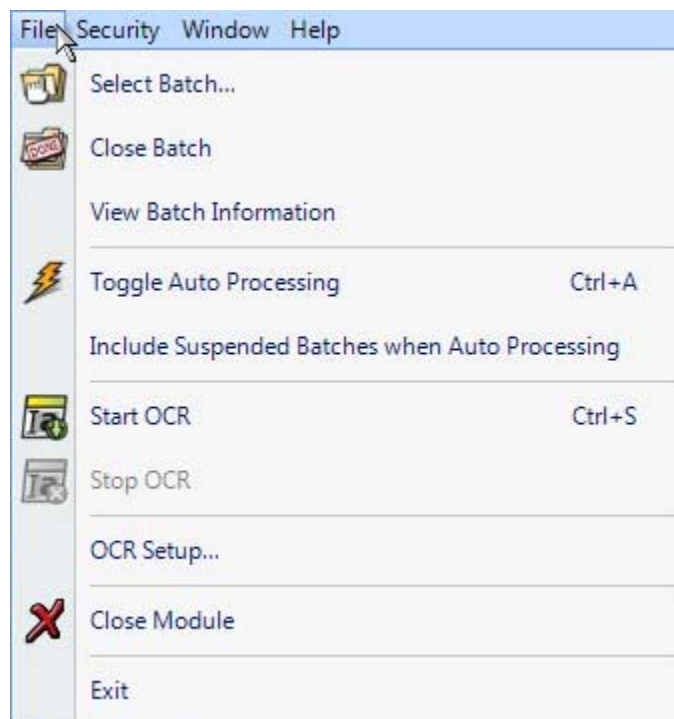
Menu Bar

Contains drop-down menus: File, Security, Window, and Help.
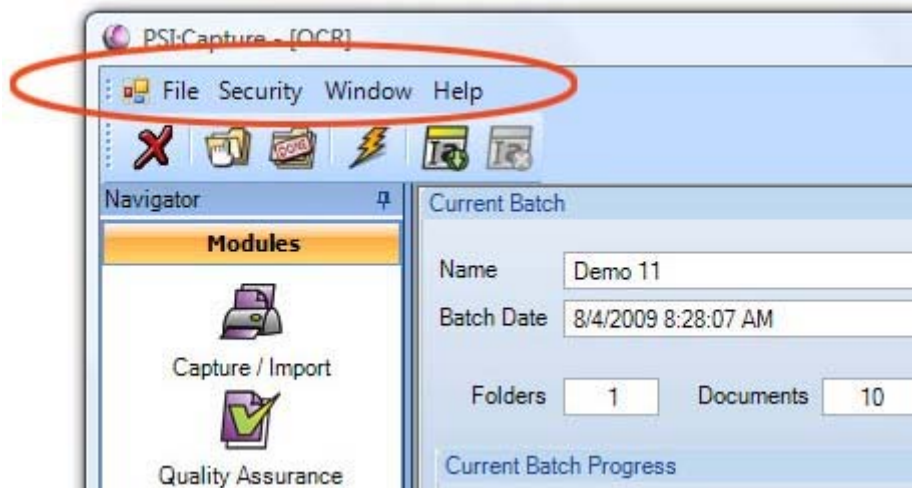
File

Contains numerous commands specific to the OCR module.

Select Batch

The "Select Batch" dialog box is brought up, and user is asked to make a selection from the list then hit "Select". All (local/shared) batches currently in the OCR workflow step are displayed here. If a batch is currently open, the user will first be prompted to close or suspend the batch.
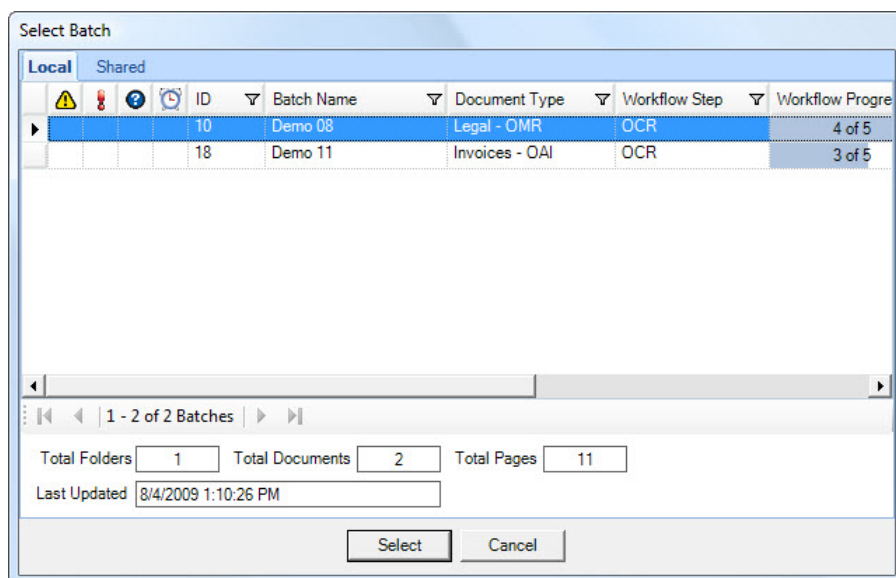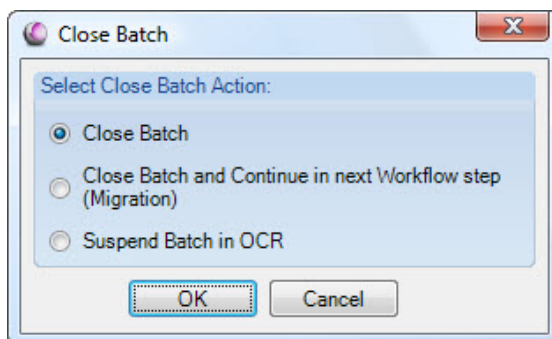

Local/Shared Tabs

A document type set as a "Local" will show all batches in the "Local" tab if it was created on that workstation. A document type set as part of a multi station deployment would typically be configured as shared batches, and therefore found in the "Shared" tab.

Close Batch

Brings up the "Close Batch" dialog box. Prompts user to select one of three options.

Close Batch

Closes the current batch, and sends the batch to the next workflow step. The user stays in the Index module, and the "Select Batch" dialog box comes up.

Close Batch and Continue in next Workflow step

Closes the current batch and the OCR module. The batch is immediately reopened in the next module. Example above shows "Migration" as the next workflow step, which is common but not always the case.

Suspend Batch in OCR

Suspends the batch in the Index workflow step for processing at a later time. The OCR Module remains open, and the "Select Batch" dialog box comes up.

View Batch Information

Opens the "Batch Information" dialog box. The button for "Modify Document Type Definition" as well as the "Batch Overrides" tab will not be available.

Main

Information on Batch Name, Batch Fields, Batch Totals, and Batch Summary are displayed. The user cannot change the batch to "High Priority" at this point, but a batch can be held by checking the "Hold Batch" box. A batch on hold will no longer be listed in the Select Batch dialog box.
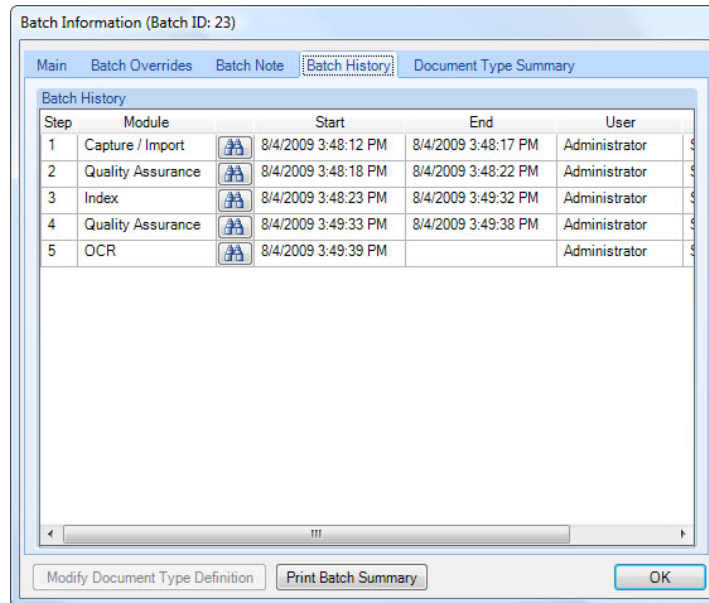
Batch Overrides

This tab is unavailable.

Batch Note

Allows the user to attach/view a note to the batch. The note will follow the batch throughout the entire workflow, and can be viewed at anytime.
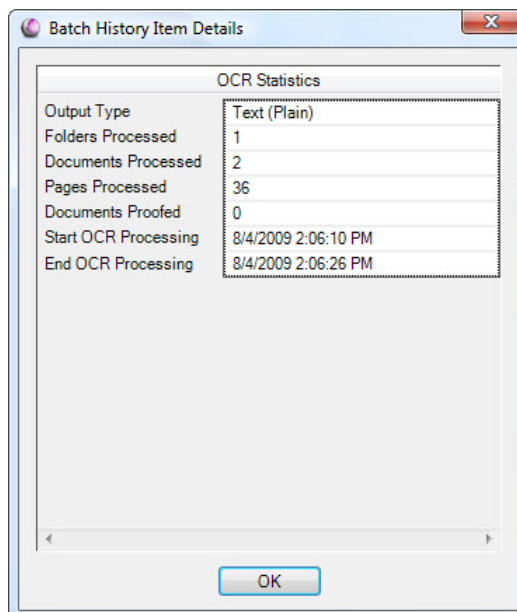
Batch History

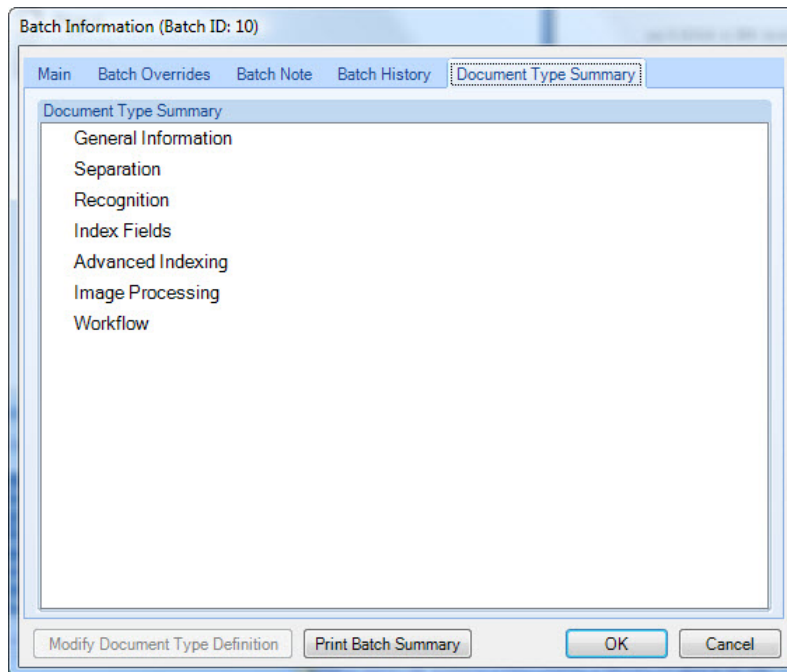Shows the history of the batch with a date timestamp.

 By clicking on the binocular icon the "Batch History Item Details" statistics on a selected workflow are shown.

Document Type Summary

A brief summary of the document type configuration settings presented in a tree structure.

Toggle Auto Processing

Auto processing is on by default when the OCR module is opened. It will wait 5 seconds before auto processing begins on any batches that are currently awaiting OCR.

Include Suspended Batches when Auto Processing

Chose whether batches suspended in the OCR workflow step can be automatically processed.

Start OCR

Begins the OCR process.

Stop OCR

Stops the OCR process.

OCR Setup

Brings up the "OCR Settings" dialog box. **NOTE:** All settings should be defined previously by the system administrator. However, some users may wish to make adjustments to the individual batch's settings to suit their needs.
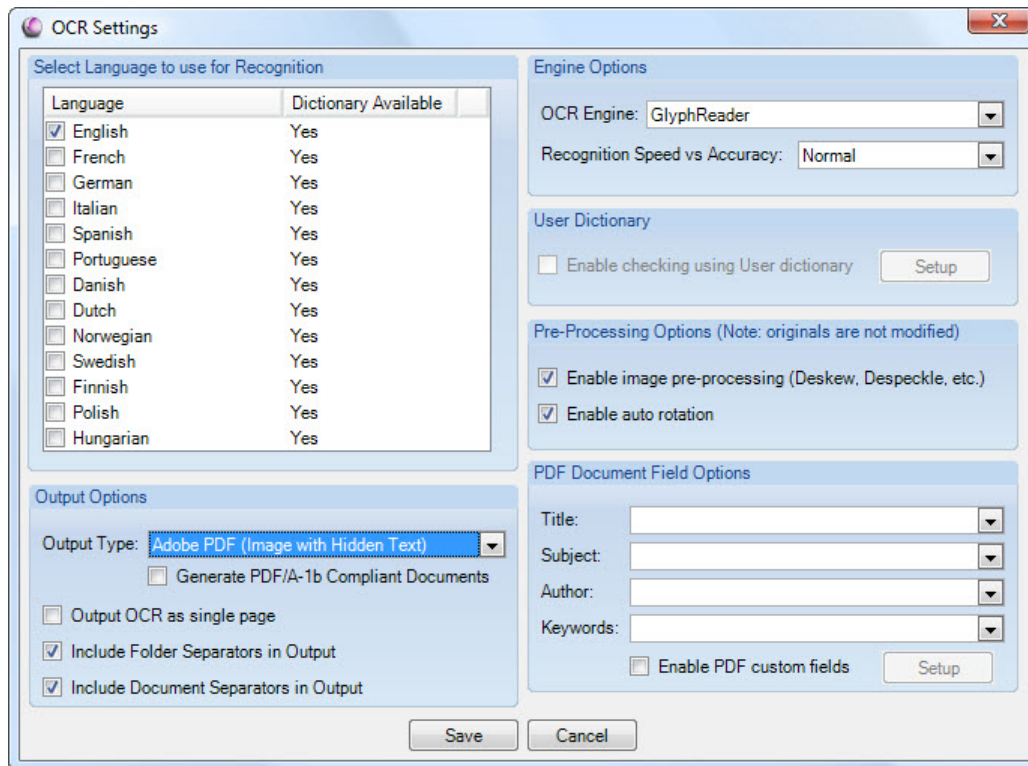
Select Language to use for Recognition

Select language(s) to perform OCR on.

Output Options

Pick one of many supported types of files to output, and choose whether to keep or delete folder/document separators.

Engine Options

Choose Enhanced, Standard, or GlyphReader as recognition engine. **NOTE:** availability of options dependent on licensed options.

User Dictionary

This option allows the user to add words to the user's own dictionary. This may be helpful when performing OCR on specialized documents, such as medical documents.

Pre-Processing Options

Enables page rotation for more accurate recognition.

PDF Document Field Options

The standard PDF Document Fields are: Title, Subject, Author and Keywords. The user can select any System, Batch, Folder or Document index field to populate the desired PDF Document Fields inside the created PDF file.

Close Module

Closes the OCR Module. **NOTE:** If a batch is currently open, user will be prompted to close or suspend the batch. PSI:Capture will remain open.

Exit

Exits out from PSI:Capture. **NOTE:** If a batch is currently open, user will be prompted to close or suspend the batch. The PSI:Capture application will then end.

Security

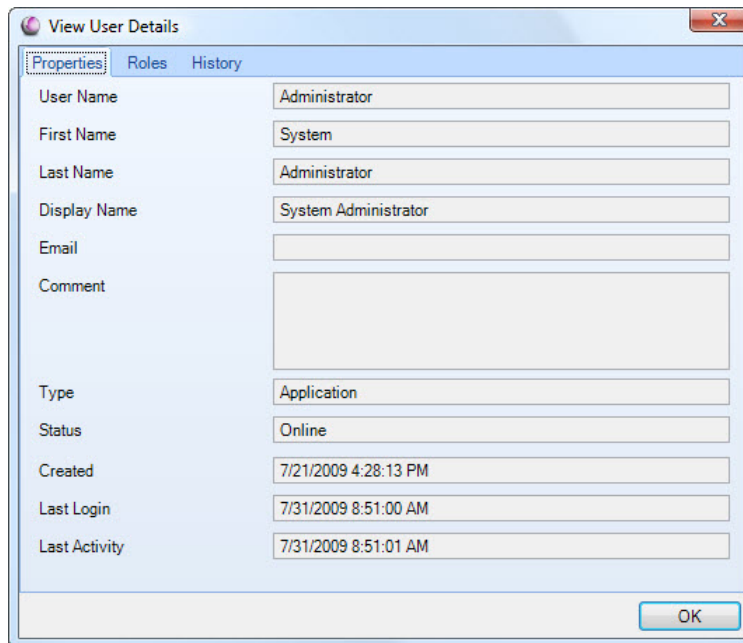Contains commands relevant to security in PSI:Capture.
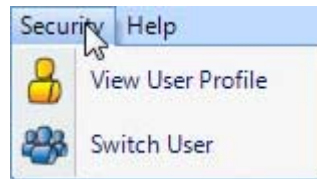
View User Profile

Opens the View User Details dialog box.

View User Details
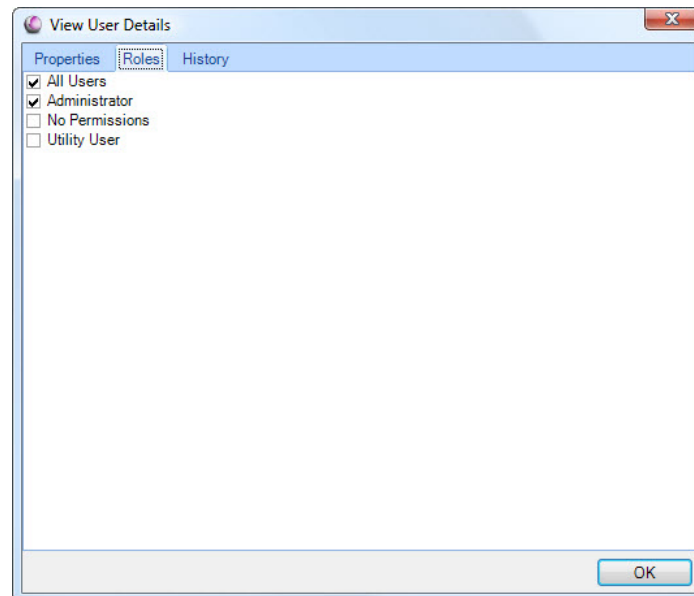
Properties
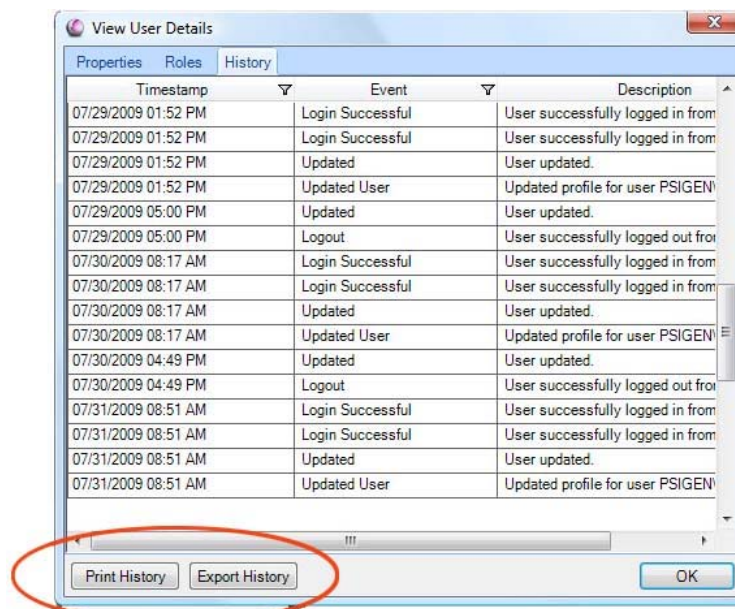
Displays details about the current user.

Roles

Displays the current user's role settings as dictated by the administrator. See "Roles" in "Security Module" in the User Manual.

History

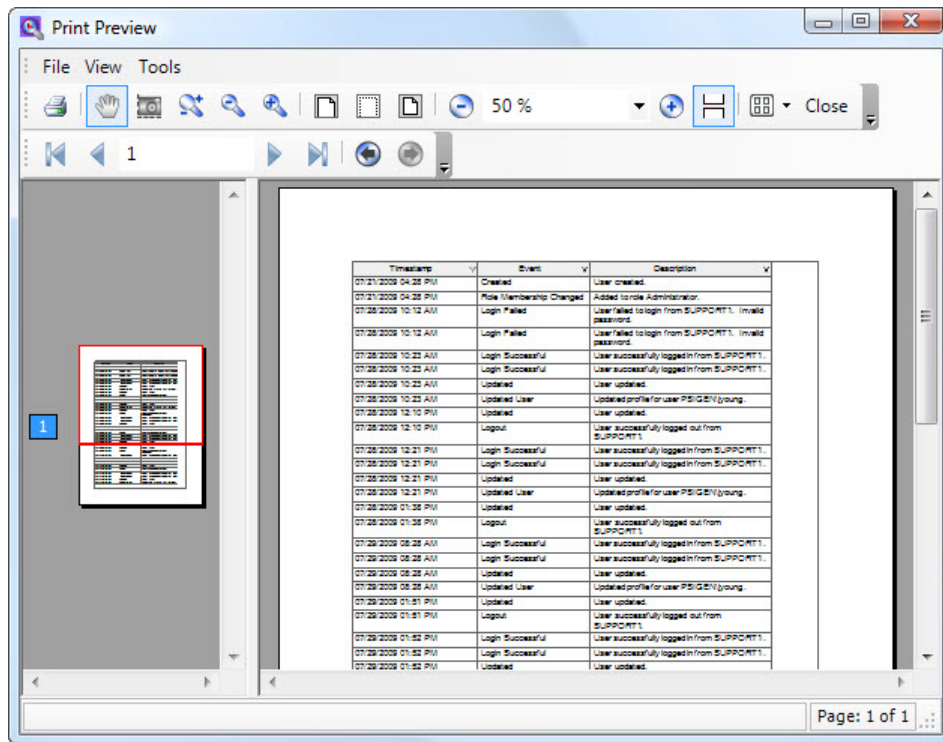Displays the history of the user's login activity in PSI:Capture.

Print History

Opens the Print Preview window.

Print Preview

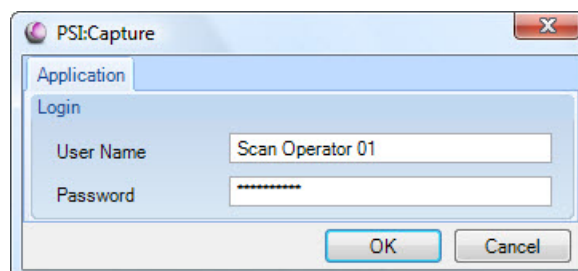Displays the standard functions for viewing and printing, such as zoom options and page setup.

Export History

Exports the login history in a Microsoft Excel Workbook file. This allows user to save the log to a desired location.

Switch User

Logs out the current user out from the PSI:Capture, and allows an alternate log in.
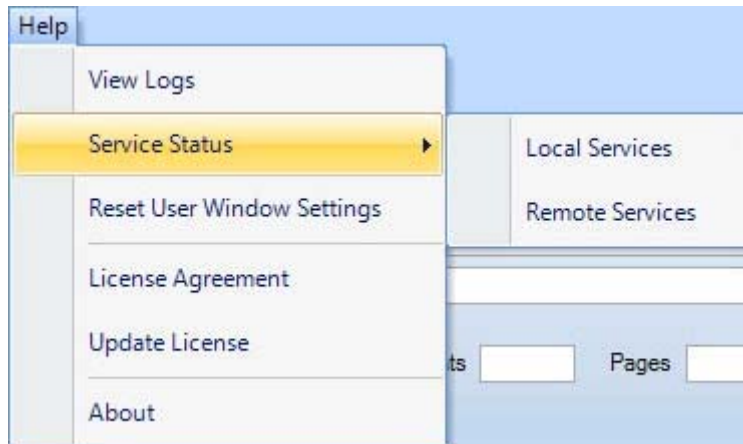
Window

This menu item is unavailable in the OCR module.

Help

Contains information and details about PSI:Capture.
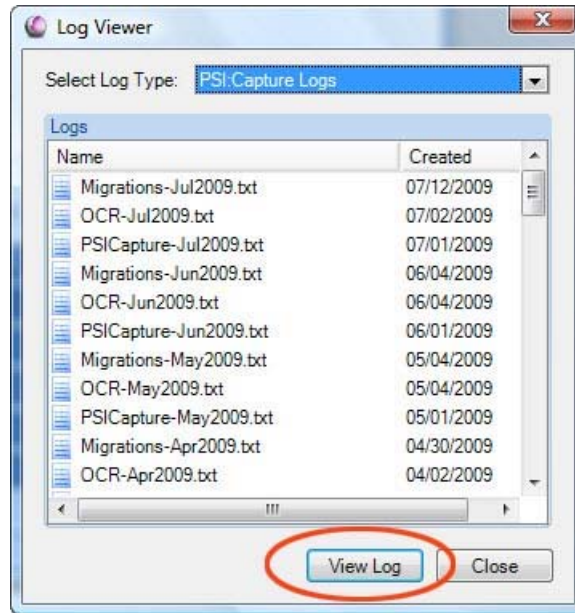

View Logs

User can select either PSI: Capture Logs or Platform Service Logs.

Log Viewer (PSI:Capture Logs)

Displays a list of logs containing all PSI:Capture modules.

View Log

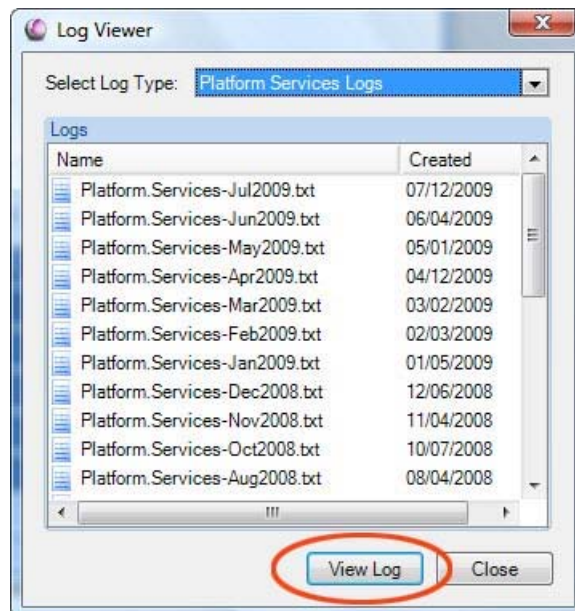Pressing this button brings up the log in your default text file viewer.

Log Viewer (Platform Services Logs)

Displays a list of logs for Platform Services.

> **View Log**

Pressing this button brings up the log in your default text file viewer.

Service Status (Local Services / Remote Services)

Shows the version and current status of the services.

Reset User Window Settings

Resets the Window settings to default.

License Agreement

Displays PSI:Capture License Agreement.

Update License

Updates PSI:Capture License with the License Server.

About

Shows PSI:Capture copyright information, maintenance expiration date, installed modules/assemblies, and license agreement.